Dose Response Designs for a Trinomial Response

Shenghua Fan and Kathryn Chaloner

School of Statistics, University of Minnesota
Technical Report #637

February, 2000

Dose Response Designs for a Trinomial Response

Shenghua Fan

Department of Statistics and Applied Probability, University of California at Santa Barbara, 5607A South Hall, Santa Barbara CA 93117, U.S.A.

email: kellyfan@pstat.ucsb.edu

and Kathryn Chaloner

School of Statistics, University of Minnesota, 313 Ford Hall,Minneapolis MN 55455, U.S.A.

SUMMARY. A dose-response relationship may be best modeled by a trinomial response: for example with the three categories "no reaction", "efficacy", and "adverse outcome". Applications of such a model in both clinical trials and toxicology will be described. Optimal designs will be found using Bayesian criteria. In addition a new concept is defined, "limiting optimality", where a sequence of designs is said to be optimal in an asymptotic sense for a sequence of prior distributions. General algebraic forms of limiting optimal

designs for a continuation-ratio model will be given. These limiting optimal designs will be shown to be very efficient and practically useful. Conditions will also be found, by empirical investigation, under which a simplifying assumption of constant slopes is reasonable in this model.

Key words: Continuation ratio; Multinomial responses; Optimal designs; Phase I/II Clinical trials.

1 Introduction

There are many important design problems where the dose response is a multinomial distribution: Zocchi and Atkinson (1999) give a toxicology example with a trinomial response to dose (survival after emergence/death after emergence/death before emergence); Thall and Russell (1998) present a new strategy for phase I/II trials where the response to dose is trinomial (no reaction/efficacy/adverse outcome); Heise and Myers (1996) discuss a clinical trial with a bivariate binary response to dose, efficacy (yes/no) and toxicity (yes/no). This last example can be though of as a multinomial response with 4 cells: the two cells corresponding to the occurrence of toxicity can be collapsed together to give a trinomial response. Other examples are given in Glonek and McCullagh (1995), Glonek (1996), and Zhu, Krewski and Ross (1994).

In this paper an ordinal trinomial response is considered and a continuation-

ratio model is adopted (Agresti, 1990, Chapter 9). The response when n_i experimental units are given a dose x_i is trinomial, (y_{1i}, y_{2i}, y_{3i}), $y_{1i}+y_{2i}+y_{3i} = n_i$. The corresponding cell probabilities are $(p_1(\theta, x_i), p_2(\theta, x_i), p_3(\theta, x_i))$, where θ is the parameters of the model. For any x and θ, $p_1(\theta, x)+p_2(\theta, x)+p_3(\theta, x) = 1$. The continuation-ratio model is

$$\log[p_3(\theta, x_i)/(1 - p_3(\theta, x_i))] = a_1 + b_1 x \quad (1)$$

$$\log[p_2(\theta, x)/p_1(\theta, x)] = a_2 + b_2 x. \quad (2)$$

For a fixed sample size n, the design problem is to choose the number of dose levels, k, the dose levels, $x_1, \ldots, x_k$, and the number of experimental units, n_i, at each x_i, such that $\Sigma n_i = n$.

To illustrate the model, a plot of the probabilities $(p_1(\theta, x), p_2(\theta, x), p_3(\theta, x))$ for $a_1 = 0, a_2 = 2, b_1 = b_2 = 1$ is shown in Figure 1. Suppose that x is the dose level, $p_1(\theta, x)$ is the probability of no reaction, $p_2(\theta, x)$ is the probability of efficacy, and $p_3(\theta, x)$ is the probability of an adverse outcome. As the dose level increases, $p_1(\theta, x)$ is increasing and $p_3(\theta, x)$ is decreasing. The probability of efficacy, $p_2(\theta, x)$, increases to a maximum and then decreases: a very low dose is likely to do nothing and an extremely high dose is likely to cause an adverse outcome or death. This is similar to Scenario 2 of Thall and Russell (1998).

An experimental design will be regarded as a probability measure on the dose domain $\mathcal{X}$. The design η puts weight m_i at dose x_i for $i = 1, 2, ..., k$,

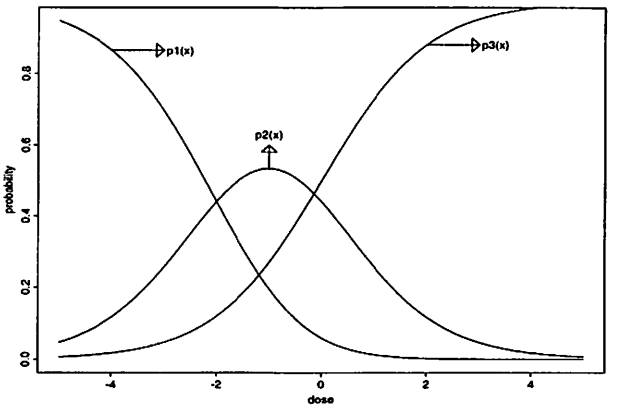


Figure 1: Probability plot: probability vs. dose for $a_1 = 0, a_2 = 2$, and $b_1 = b_2 = 1$.

$m_i = n_i/n$ the fraction of observations to be taken at that dose. The values $m_i, i = 1, \ldots, k$, are non-negative and sum to one over i, but are not otherwise constrained. A design can therefore be thought of as a probability measure, η on $\mathcal{X}$. A design putting weight m_i at dose x_i for $i = 1, 2, ..., k$ will be written as $(m_1, m_2, ..., m_k)$ at $(x_1, x_2, ..., x_k)$. For a given sample size n the values nm_i's can be rounded to integers in a systematic way to run an experiment (Pukelsheim, 1993, Chapter 12).

Designs are derived for the model described by (1) and (2) which optimize several criteria based on the Fisher information matrix (see Silvey, 1980). The inverse of the Fisher information matrix gives the asymptotic variance covariance matrix for the maximum likelihood estimate (MLE). The information matrix depends on both the design, denoted by η, and the parameter values, denoted by θ. Expressions for the information matrix, $M(\theta, \eta)$, of the constant-slope model, where $b_1 = b_2$, and also of the more general model

where the slopes b_1 and b_2 are not necessarily equal, are in Appendix A.

Local D-optimality chooses the design which maximizes the determinant of the information matrix at a single value of the parameters; that is, it is the η that maximizes $\phi(\eta) = \text{logdet}M(\theta, \eta)$. This criterion should make the asymptotic variances of the MLE's small, in a general sense, if the specified value of θ is close to the true value of θ. The same criterion averaged over a distribution on the parameters, π, gives "Bayesian" D-optimality: $\phi(\eta) = E_\pi \text{logdet}M(\theta, \eta)$. This criterion is typically a more robust criterion than local D-optimality and is better over a range of values for θ. An alternative criterion, c-optimality, will also be used, which minimizes the asymptotic variance of a specific quantity of interest. Suppose that the quantity of interest is a single function of the parameters, denoted $g(\theta)$, and denote the gradient vector of $g(\theta)$ as $c(\theta) = \frac{\partial g(\theta)}{\partial \theta}$ (as used in the delta method for calculating an asymptotic variance, see eg. Agresti, 1990, Chapter 12). Then the c-optimality criterion is $\phi(\eta) = -E_\pi c(\theta)^T M(\theta, \eta)^{-1} c(\theta)$. The expectation is either over a prior distribution π (Bayesian c-optimality) or over a point mass distribution (local c-optimality). The specific example examined here is where $g(\theta) = x_{max}$, the dose at which the probability of efficacy is maximized. Putting a distribution on the unknown parameters to reflect the uncertainty prior to experimentation is very appealing from a design perspective: see Verdinelli (1992) and Chaloner and Verdinelli (1995) for a review of Bayesian approaches to design. Applications and examples of locally optimal designs are given in Kitsos, Titterington, and Torsney (1988), Ford, Torsney, and

Wu (1992), and Wu (1988). For a general discussion of the design problem for dose response see Wong and Lachenbruch (1996).

1.1 Examples

Example 1: Toxicological Example – The following example is reported in Zocchi and Atkinson, together with a data set. In the data set $n = 3,500$ house fly pupae (*Musca domestica* L., 1758) are assigned to be exposed to a level, x_i, of gamma radiation, with $n_i = 500$ pupae at each level $i = 1, \ldots, 7$. After a period of time the number of flies that emerge from the pupae and survive is denoted y_1, the number that died after emergence is denoted y_2, and the number that died before emergence is denoted y_3. (Zocchi and Atkinson used different notation and a different model but the model, (1) and (2), is similar.) The design problem is to choose the number of levels and doses of gamma radiation, and to choose the number of pupae to assign each dose.

In this example sequential design is not an option because a period of time elapses between the application of the dose, x_i, and the response, y_i. Batch sequential design might be an option (as in Sitter and Wu, 1999) but is not explicitly considered here. Ethical considerations do not enter into this design example as they do in Example 2.

Example 2: Phase I/II Clinical Trial – Thall and Russell (1998) propose a practical new strategy for phase I/II trial design and conduct. They describe several such phase I/II trials. One is a cancer clinical trial where patients with advanced hematologic malignancies or lymphoma are assigned

to a combination drug dose x. Two response variables are of interest: severe toxicity (yes/no) and graft-versus-host disease (none, moderate, severe). The response is therefore a six cell multinomial. As the adverse outcome is very severe, and includes death, Thall and Russell collapse into 3 cells of interest (no reaction/efficacy/adverse outcome). Because it is important not to expose patients to undue toxicity, important dose limiting constraints enter into the design. Thall and Russell use a sequential strategy, for finding a dose satisfying both safety and efficacy requirements, based on Bayesian methods. As the designs derived in this paper are non-sequential, and do not take ethical constraints into account, they are not immediately applicable to this particular class of clinical trials, but they do provide a useful benchmark with which to compare, and possibly improve upon, the designs of Thall and Russell. Additional discussion of ethical constraints is given in Thall, Estey, and Sung (1999) together with a description of an ongoing trial for donor lymphocyte infusion which uses their strategy.

1.2 Summary

A new concept of "limiting optimality" is introduced in Section 2 for a sequence of designs which are approximately optimal. This concept enables the construction of designs with closed form expressions. Closed form designs provide starting points for numerical algorithms and they also provide benchmarks against which sequential, or other strategies can be evaluated. They are also shown to be remarkably efficient for the models studied here.

Section 3 gives designs for the constant-slope model ($b_1 = b_2$), and also for the model where the slopes are not necessarily equal. Designs for Example 2 are given in Section 4. Finally, Thall and Russell (1998) use a three parameter model making an assumption that the slopes are equal: this assumption is investigated empirically in Section 5.

2 Limiting Optimal Designs

Designs are thought of as probability measures and the criteria to be used are all concave functions over the set of probability measures, $\mathcal{H}$, on the design region $\mathcal{X}$ (see Silvey, 1980). The General Equivalence Theorem (as given in Whittle, 1973, and extended in Chaloner and Larntz, 1989) can be used and is given in Appendix B. The optimization is multidimensional: the number and values of support points and the weight at each support point must be chosen. The theorem means that after a candidate design has been found, by numerical or other methods, its global optimality can be easily verified by merely checking that a simple function of the variable x is non-positive. Define a one point design, putting all mass at $x \in \mathcal{X}$ as η_x. Then a design η_0 at which the criterion, ϕ, is differentiable is optimal for a prior distribution π, if, and only if, the directional derivative $F_\phi(\eta_0, \eta_x, \pi)$ at η_0 in the direction of all one-point designs, η_x for x in $\mathcal{X}$, is non-positive. (If the distribution π is a one-point probability measure then the criterion is that of local optimality.) The directional derivative at η_0 in the direction η for a prior distribution π

is:

$$F_\phi(\eta_0, \eta, \pi) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}[E_\pi[\phi\{(1-\epsilon)\eta_0 + \epsilon\eta\} - \phi\{\eta_0\}]].$$

For D-optimality and c-optimality, $F_\phi(\eta_0, \eta, \pi)$ is calculated easily: formulas are given in Appendix B.

The following defines a new concept of a sequence of limiting optimal designs.

Definition 1. For a concave criterion ϕ on a set of design measures $\mathcal{H}$, a sequence of designs, $\{\eta_i, i \text{ is an integer}\}$, is called a sequence of limiting optimal designs ϕ for a sequence of prior distributions, $\{\pi_i, i \text{ is an integer}\}$, if

$$\sup_{\eta \in \mathcal{H}} F_\phi(\eta_i, \eta, \pi_i) = d_i > 0$$

and $d_i \to 0$ as $i \to \infty$. Each design of this sequence, say η_i, is called a limiting optimal design of π_i.

These limiting optimal designs are not optimal but tend to be optimal as the index i goes to infinity.

Lemma 2.1. *Consider a concave criterion ϕ, a sequence of priors, π_i, and the corresponding limiting optimal designs, η_i. Let η_i^* be the Bayesian optimal design for prior π_i. If the limit and integral in $F_\phi(\eta_i, \eta_i^*, \pi_i)$ are reversible, i.e. $F_\phi(\eta_i, \eta_i^*, \pi_i) = E_{\pi_i}[\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\{\phi(\theta, (1-\epsilon)\eta_i + \epsilon\eta_i^*) - \phi(\theta, \eta_i)\}]$, then,*

$\phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i)$, the difference in the value of the criterion at design η_i^*, and at design η_i, goes down to zero as $i \to \infty$.

Proof. See Appendix C. □

If a criterion ϕ is concave and differentiable at η_i for each integer i, then $\{\eta_i, i \text{ is an integer}\}$ is a sequence of limiting optimal designs for a concave criterion ϕ for a sequence of prior distributions, $\{\pi_i, i \text{ is an integer}\}$, if

$$\sup_{x \in \mathcal{X}} F_\phi(\eta_i, \eta_x, \pi_i) = d_i > 0$$

and $d_i \to 0$ as $i \to \infty$. Because these limiting designs are not exactly optimal, it is important to examine their efficiencies. The following definition of efficiency is used.

Definition 2. The efficiency of a design η is defined to be the sample size required for an experiment using the optimal design to reach the same value of the criterion as an experiment using design η with sample size one.

3 Optimal Designs

Constant slopes ($b_1 = b_2$) will be initially assumed. The three cells of the trinomial response are denoted as no reaction/efficacy/adverse outcome. Let $u = a_2 - a_1$. Figure 2 indicates that the value of u is a measure of how good the drug/therapy is. Larger values of u give wider ranges of dose level with low probability of adverse outcome in the range of efficacy. A negative

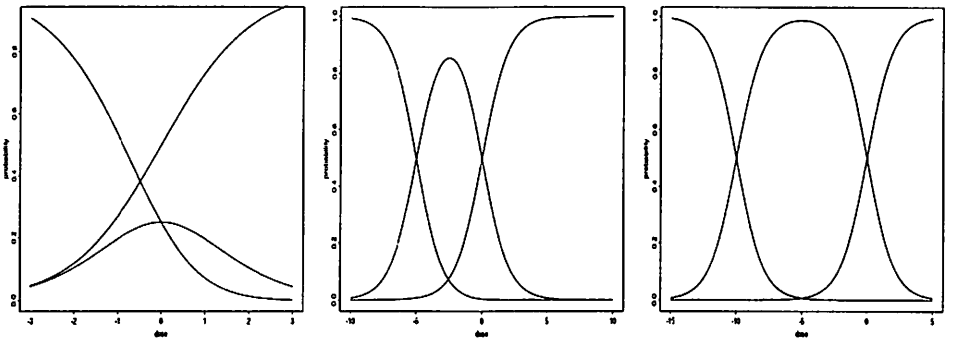


Figure 2: From the left to the right: probability plot, probability vs. dose, for $a_1 = 0, b_1 = 1$, and $u = 0, 5$, and 10, respectively.

u value indicates a bad drug since the probability of efficacy is very low at most dose levels. Negative u values will, therefore, be only briefly discussed. For convenience θ will be rewritten as (u, a_1, b_1) but the information matrix with $\theta = (a_1, b_1, a_2)$ in Appendix A is used to find the optimal designs.

D-optimal Designs

Define two sets of parameters: $\theta_0 = (u, 0, 1)$, and $\theta = (u, a_1, b_1)$. Also define design $\eta_0 = (m_1, m_2, ..., m_k)$ at $(x_1, x_2, ..., x_k)$ and design $\eta = (m_1, m_2, ..., m_k)$ at $(\frac{x_1-a_1}{b_1}, \frac{x_2-a_1}{b_1}, ..., \frac{x_k-a_1}{b_1})$. Then it is easily shown algebraically that: $\det(M(\theta, \eta)) = b_1^{-2} \det(M(\theta_0, \eta_0))$. Therefore

Lemma 3.1. *If* $\eta_0^* = (m_1^*, m_2^*, ..., m_k^*)$ *at* $(x_1^*, x_2^*, ..., x_k^*)$ *is locally D-optimal for* $\theta_0 = (u, 0, 1)$ *then* $\eta^* = (m_1^*, m_2^*, ..., m_k^*)$ *at* $(\frac{x_1^*-a_1}{b_1}, \frac{x_2^*-a_1}{b_1}, ..., \frac{x_k^*-a_1}{b_1})$ *is also locally D-optimal for* $\theta = (u, a_1, b_1)$.

Proof. As $\det(M(\theta, \eta)) = b_1^{-2} \det(M(\theta_0, \eta_0))$ and the transformation $\eta_0 \mapsto \eta$

is one-to-one and onto, maximizing $\log\det(M(\theta,\eta))$ is the same as maximizing $\log\det(M(\theta_0,\eta_0))$ over all possible designs η and η_0. □

Without loss of generality, therefore, $a_1 = 0, b_1 = 1$ will be assumed for deriving locally D-optimal designs.

Locally D-optimal designs are first found numerically. The number of support points of the locally D-optimal design depends on the value of $u = a_2 - a_1$ alone. If u is positive then the locally D-optimal design has either two or three or four support points. Figure 3 illustrates that for a small u value, $u = 0$ say, the locally D-optimal design is 2-point; for a slightly larger u value, $u = 5$, it is 3-point; and for a large u value, $u = 10$, it is 4-point. The numerical optimization guarantees a local optimum, and global optimality can be checked using the General Equivalence Theorem in Appendix B which just requires that its directional derivative is everywhere non-positive. Figure 4 shows the directional derivatives, and they are nonpositive and equal to zero at the design points. If u is negative, the locally D-optimal designs are typically 2-point designs.

The probability plot for $u = 10$ in Figure 2 looks like two separate, single logistic regressions in separate regions of the dose range. One logistic regression is on the left and is for no reaction/efficacy, with adverse outcome having a negligible probability, and the other is on the right for efficacy/adverse outcome, with the probability of no reaction being negligible. >From results for a single logistic regression (White, 1975, for example), a 4-point design with equal weights, two points for each logistic regression, seems therefore a good

initial guess. The following result shows that such a strategy is, in fact, limiting D-optimal.

Theorem 3.2. *Suppose* $a_1 = 0$ *and* $b_1 = 1$. *As* u *goes to* ∞, *the design* η^* *putting equal weight at* $x = \{1.223, -1.223, -u + 1.223, -u - 1.223\}$ *is limiting locally D-optimal.*

Proof. See Fan (1999, Chapter 5). The proof shows that let u be the index i and then d_i, as in Definition 1, is such that $d_i \to 0$ as $i \to \infty$. □

The probability of the responses at the four design points are shown in Table 1. In addition, the limits as $u \to \infty$, of these probabilities are given in Table 2.

x (dose)	no reaction	efficacy	adverse outcome
1.223	$0.2274\frac{1}{1+e^{u+1.223}}$	$0.2274\frac{e^{u+1.223}}{1+e^{u+1.223}}$	0.7726
-1.223	$0.7726\frac{1}{1+e^{u-1.223}}$	$0.7726\frac{e^{u-1.223}}{1+e^{u-1.223}}$	0.2274
$-u+1.223$	$0.2274\frac{1}{1+e^{-u+1.223}}$	$0.7726\frac{1}{1+e^{-u+1.223}}$	$\frac{e^{-u+1.223}}{1+e^{-u+1.223}}$
$-u-1.223$	$0.7726\frac{1}{1+e^{-u-1.223}}$	$0.2274\frac{1}{1+e^{-u-1.223}}$	$\frac{e^{-u-1.223}}{1+e^{-u-1.223}}$

Table 1: Probabilities at design points of the limiting design.

x (dose)	no reaction	efficacy	adverse outcome
1.223	0	0.2274	0.7726
-1.223	0	0.7726	0.2274
$-u+1.223$	0.2274	0.7726	0
$-u-1.223$	0.7726	0.2274	0

Table 2: The limits of probabilities at design points of the limiting design, as u goes to ∞.

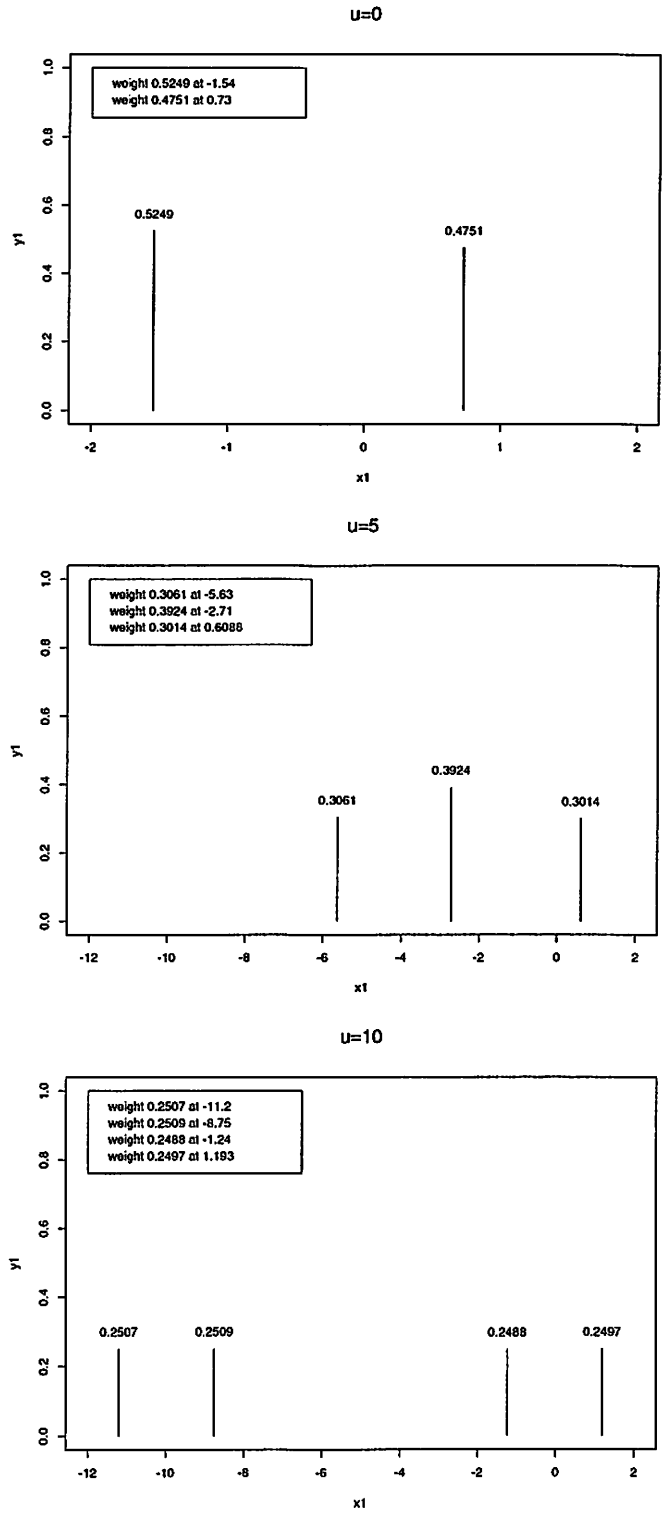
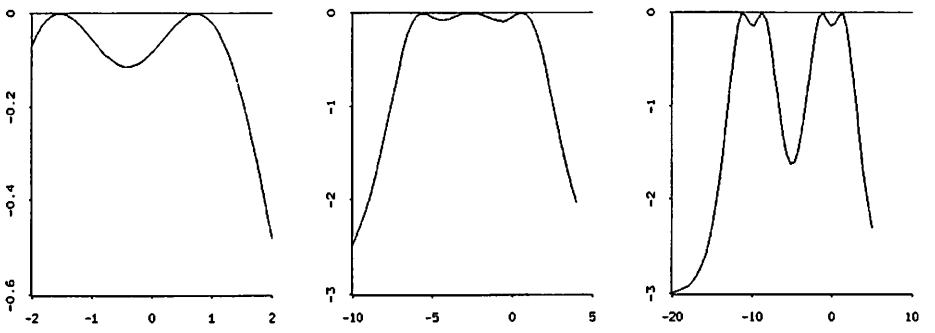


Figure 3: Locally D-optimal designs for $u = 0, 5$, and 10.



(a) $u = 0$ (b) $u = 5$ (c) $u = 10$

Figure 4: Directional derivative plots of D-optimal designs for $a_1 = 0$, $b_1 = 1$, and several u values.

Lemma 2.1 guarantees that for large u values these limiting designs are efficient. Figure 5 shows that the efficiencies (as defined in Definition 2) for small u values are suprizingly efficient: efficiencies are all higher than 97%.

As u goes to $-\infty$, $p_2(\theta, x)$ tends to 0, so this model becomes that of a single logistic regression. A candidate limiting locally D-optimal design therefore puts weight $1/2$ at each of $a_1 \pm 1.543 b_1$, the well documented locally optimal design for a single logistic regression. The efficiency of this design for $u = -20$ is 95.7%.

A general form of Bayesian D-optimal designs, for arbitrary prior distributions, is intractable in closed form but designs can be found numerically (as in Section 4). For any prior distribution putting uncertainty only on u (for fixed a_1 and b_1), a general property of Bayesian D-optimal designs is given here. In addition, for a simple prior distribution, a general form of (limiting)

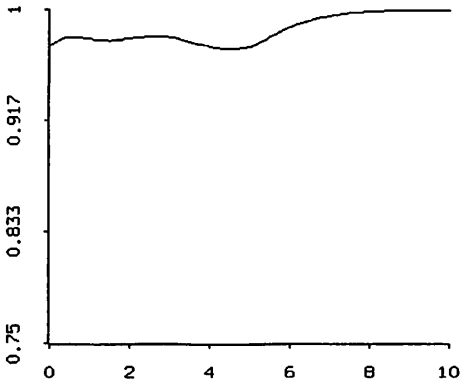


Figure 5: Efficiency plot of limiting locally D-optimal design: efficiency vs u.

Bayesian D-optimal designs is found and is shown to be very efficient.

Let the prior distribution π put weight π_i at $\theta = (u_i, a_1, b_1)$ and let the prior distribution π_0 put weight π_i at $\theta = (u_i, 0, 1)$. It can be shown by straightforward algebra that if $\eta_0^* = (m_1^*, m_2^*, ..., m_k^*)$ at $(x_1^*, x_2^*, ..., x_k^*)$ is the Bayesian D-optimal design for prior π_0 then $\eta^* = (m_1^*, m_2^*, ..., m_k^*)$ at $(\frac{x_1^*-a_1}{b_1}, \frac{x_2^*-a_1}{b_1}, ..., \frac{x_k^*-a_1}{b_1})$ is the Bayesian D-optimal design for prior π.

Without loss of generality, therefore, again assume $a_1 = 0, b_1 = 1$. Consider a simple prior distribution π with two equally weighted support points: $(0, 0, 1)$ and $(u, 0, 1)$. The following result was suggested by numerical exploration and was then proved.

Theorem 3.3. *Let the prior π be as above. Then the design η putting weight* $0.16667, 0.44825, 0.38508$ *at* $-u, -1.47128, 1.14271$, *respectively, is a limiting Bayesian D-optimal design for the prior distribution π as $u \to \infty$.*

Proof. See Fan (1999, Chapter 5) for a proof similar to that of Theorem

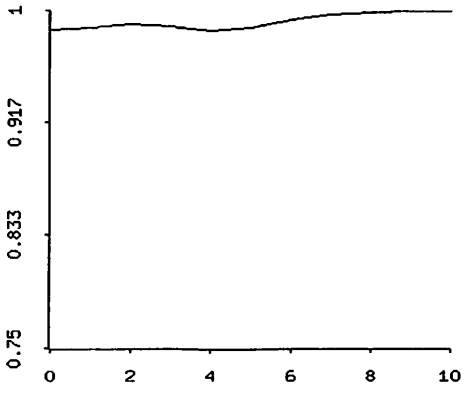


Figure 6: Efficiency plot of limiting Bayesian D-optimal design: efficiency vs. u

3.2. □

The limiting designs are again very efficient, as is shown in Figure 6: all efficiencies, as defined in Definition 2, are higher than 97.8%.

For unequally weighted prior distributions, Fan (1999, Chapter 5) gives some examples and conjectures for the limiting Bayesian optimal design.

c-optimal Designs

Suppose the goal of the experiment is finding the dose, x_{max}, which maximizes $p_2(\theta, x)$, the probability of efficacy. The value $x_{max} = -\frac{a_1+a_2}{2b_1}$ is a function of θ: $x_{max} = g(\theta)$. Let $c(\theta)$ be the gradient vector of $g(\theta)$. The asymptotic variance of the MLE of x_{max}, for a design η, is $c(\theta)^T M(\theta, \eta)^{-1} c(\theta)$. Then c-optimality minimizes this asymptotic variance.

Similarly to Lemma 3.1, if $\eta_0^* = (m_1^*, m_2^*, ..., m_k^*)$ at $(x_1^*, x_2^*, ..., x_k^*)$ is locally

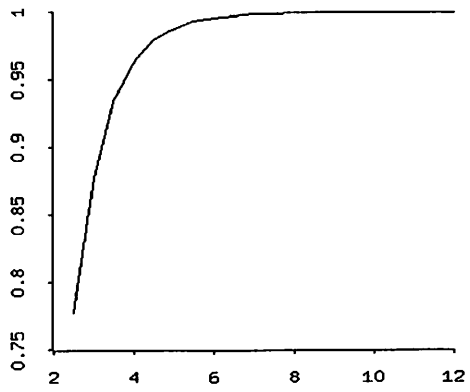


Figure 7: Efficiency plot of limiting c-optimal designs: efficiency vs. u.

c-optimal for $\theta_0 = (u, 0, 1)$ then $\eta^* = (m_1^*, m_2^*, ..., m_k^*)$ at $(\frac{x_1^*-a_1}{b_1}, \frac{x_2^*-a_1}{b_1}, ..., \frac{x_k^*-a_1}{b_1})$ is also locally c-optimal for $\theta = (u, a_1, b_1)$. Without loss of generality, therefore, only c-optimal designs for $a_1 = 0, b_1 = 1$ will be explored here.

Theorem 3.4. *Suppose $a_1 = 0, b_1 = 1$. As $u = a_2 - a_1$ goes to ∞, the design η putting $1/2$ weight at each of $x = 0$ and $x = -u$ is a limiting locally c-optimal design for x_{max}.*

Proof. See Fan (1999, Chapter 5). □

For u between 2.5 and 4 these limiting designs are over 75% efficient, and for u greater than 4 over 95% efficient, as shown in Figure 7.

For a very small value of u, 0.1 say, numerical problems sometimes arise finding locally c-optimal designs. The limiting designs above can serve as starting designs for optimization to avoid such problems. When u is 0 the c-optimal design becomes a singular one-point design.

Theorem 3.5. *The c-optimal design of* $a_1 = 0, b_1 = 1$, *and* $u = 0$ *is a one-point design putting mass* 1 *at* $x = 0$.

Proof. See Fan (1999, Chapter 5) for the proof which uses an equivalence theorem for singular optimal designs, as given, for example, in Silvey (1980, Chapter 3). □

Singular designs are of extremely limited direct practical use as the parameters are not all estimable. They are however useful in sequential strategies and as benchmarks.

Different Slopes

For a more general model where b_1 is not necessarily equal to b_2, limiting locally D-optimal and Bayesian D-optimal designs are given in Fan (1999, Chapter 6). They are, however, not as efficient as in the constant-slope model. In some cases, the efficiency can be as little as 60%, even for quite large values of u, $u = 20$ say. The criterion for c-optimality for estimating x_{max} is not straightforward to implement in this more general model as there is no closed form expression for x_{max}.

4 Optimal Designs for Example 2

Thall and Russell (1998) provided a prior distribution determined by clinicians and statisticians together. A discrete approximation to their prior

distribution, which simplifies the numerical optimization, together with our slightly different model, is used here for illustration. Only three support points for each parameter are used, centered at the values used by Thall and Russell: $-6, -3.5, -1$ for a_1; $-5, -1, 3$ for a_2; $0.04, 0.22, 0.40$ for b_1. The prior distribution has $3 \times 3 \times 3 = 27$ support points and is uniform, with weight $1/27$ at each point. The Bayesian D-optimal design is found numerically, and is an 8-point design, putting weight $0.0173, 0.1143, 0.2049, 0.1736, 0.1904, 0.1288, 0.06504$, and 0.1058 at $x = -94, -8.97, -0.185, 6.064, 11.35, 17.54, 28.08$, and 124, respectively. Its optimality has been verified by its directional derivative plot.

The Bayesian c-optimal design for estimating the dose with the highest probability of efficacy, x_{max}, is also found numerically. It is also an 8-point design, putting weight 0.0290868, 0.128426, 0.152321, 0.276159, 0.103444, 0.101994, 0.156102, 0.0524675 at $x = -86.2186, -8.9973$, 2.61665, 9.84083, 17.5687, 25.9459, 107.829, and 142.552, respectively. Its optimality has also been verified by its directional derivative plot.

5 When Can Constant Slopes be Assumed?

Thall and Russell (1998) assume constant slopes for their very similar model, but they give no evidence justifying this assumption. The robustness of constant slopes assumption will be explored here in the following three ways:

1. examining the efficiency of the optimal design for the constant-slope

model when the slopes are actually not equal.

2. examining the values to which the maximum likelihood estimates of the constant-slope model converge, and by how much the fitted probability curves differ from the true probability curves, when the slopes are assumed equal but are not.

3. examining the performance of the optimal design for the constant-slope model for a small sample when the slopes are actually not equal.

The parameter vector of the different-slope model is denoted as θ and that of the constant-slope model is denoted as $\widetilde{\theta} = (u, a_1, b_1)$.

5.1 Efficiency of Optimal Designs

Let $r = b_1/b_2$ and $d = a_1 - ra_2$. From Equation (1) and (2), it can be shown that for fixed a_2, b_2, and r, the smaller d is, the better this drug/therapy is (the higher the probability of efficacy and the lower the probability of adverse outcome). This can also be seen from Figure 8, which shows probability plots for $a_2 = 0, b_2 = 1$, $r = 0.8$, and $d = -5, 0$.

Let $\widetilde{\eta}^*$ be the locally D-optimal design for the constant-slope model with $\widetilde{\theta} = (a_2 - a_1, a_1, b_1)$, and η^* be the locally D-optimal design of the model with θ having the same values of a_1, a_2, and b_1, but b_2 not necessarily equal to b_1: termed the different-slope model.

Lemma 5.1. *For fixed a_1, a_2 and r, under D-optimality and the different-slope model, the efficiency of $\widetilde{\eta}^*$ (compared to η^*, as in Definition 2) does not*

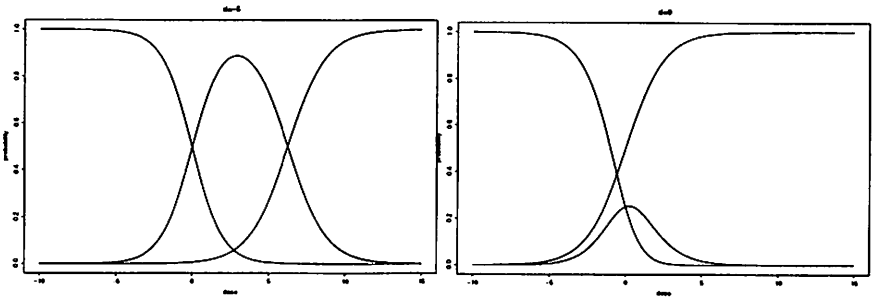


Figure 8: Probability plots, probability vs. dose, for $a_2 = 0, b_2 = 1, r = 0.8$, and $d = -5$ (left); $d = 0$ (right).

depend on the value of b_2.

Without loss of generality $b_2 = 1$ is assumed for this subsection.

The efficiency of $\widetilde{\eta}^*$, the locally D-optimal design of the constant-slope model, under the different-slope model were studied in Fan (1999, Chapter 7). In this dissertation efficiency plots of efficiency against d were drawn for $a_2 = -1, 0, 1, 2, 5$, and $r = 0.5, 0.8, 0.9, 1.0, 1.1, 1.2, 2.0$. They were all similar and the plot for $a_2 = -1$ is shown in Figure 9 for illustration. Overall, for this range of parameter values, if $0.9 \leq r \leq 1.1$, that is b_1 is within 10% of b_2, the efficiency is over, or very close to, 90%.

5.2 Convergence of MLE's

Huber (1967) showed that if the true distribution does not lie within the specified parametric class ($b_1 = b_2$ here) then, as the sample size goes to

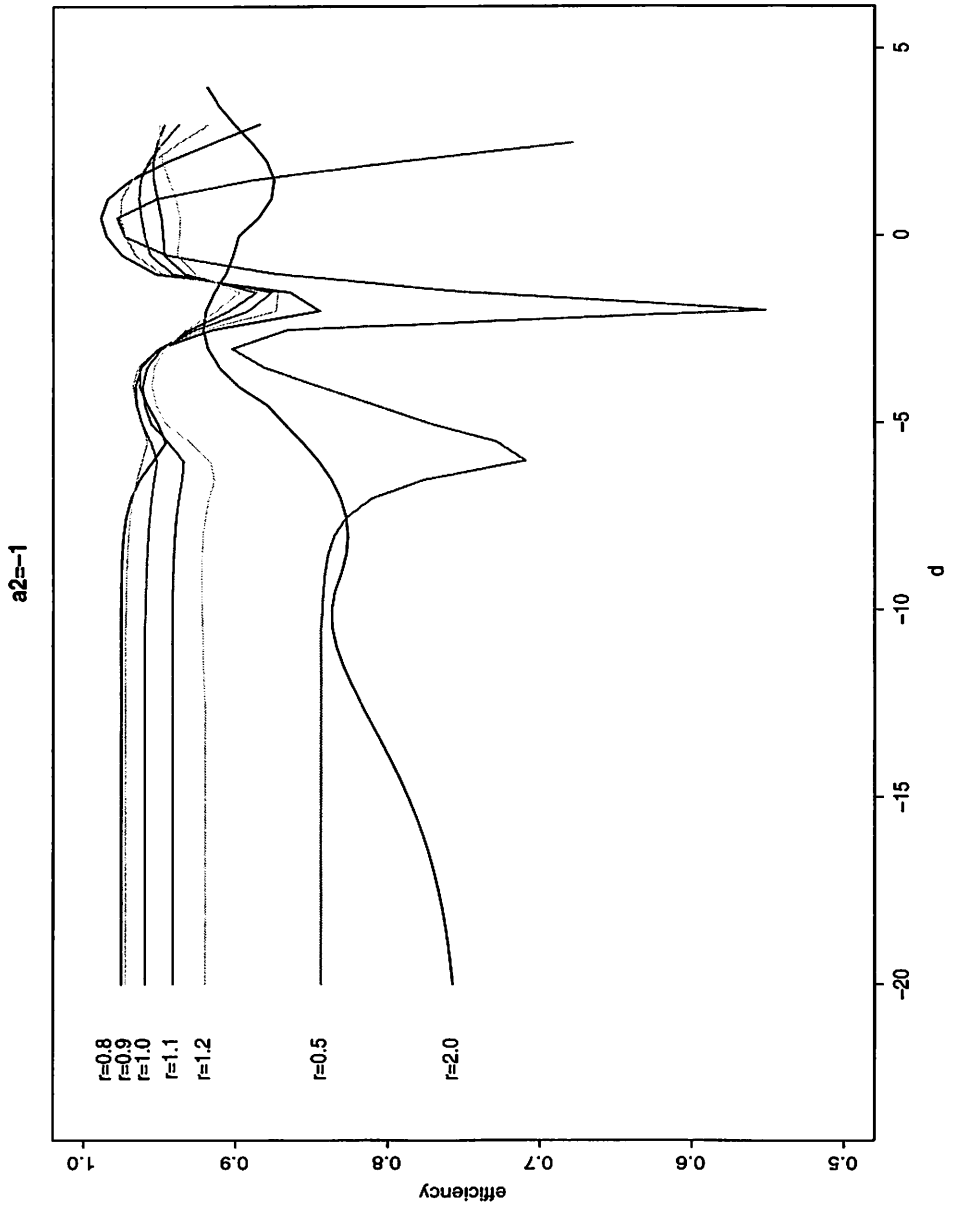


Figure 9: Efficiency plot for $a_2 = -1$: efficiency vs. d.

∞, the MLE approaches the value corresponding to the distribution in the parametric family closest to the true distribution in terms of the directed Kullback divergence. For any design, therefore, these values can be calculated numerically.

Suppose that the constant-slope model is adopted for data analysis and the D-optimal (or c-optimal) designs for the constant-slope model are used.

Lemma 5.2. *Let the true parameter vector in the different-slope model, θ, be $(a_1, b_1, a_2, b_2 = b_1/r)$. If the MLE in the constant-slope model, $(\widehat{a_1}, \widehat{b_1}, \widehat{a_2})$, converges to (a_1^*, b_1^*, a_2^*) for $b_1 = 1$, then for $b_1 = b$ and the same a_1, a_2 and r values, this MLE converges to (a_1^*, bb_1^*, a_2^*).*

Without loss of generality $b_1 = 1$ will be assumed for this subsection.

Using D-optimal Designs for the Constant-Slope Model

The true probability curves and the asymptotic fitted probability curves were drawn on the same plot for 208 parameter values. The 208 plots drawn corresponded to: $a_2 - a_1 = -5, 0, 5, 10$, $a_1 = -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5$, $b_1 = 1$, and $b_2 = 0.9, 0.95, 1.05, 1.1$. In the plots where b_2 is less than 10% different from b_1, the asymptotic fitted probability curves can barely be distinguished from the true probability curves. Figure 10 is an example.

Using c-optimal Designs for the Constant-Slope Model

This part will study the performance of c-optimal designs for the constant-slope model. The error of both estimates and confidence intervals used for

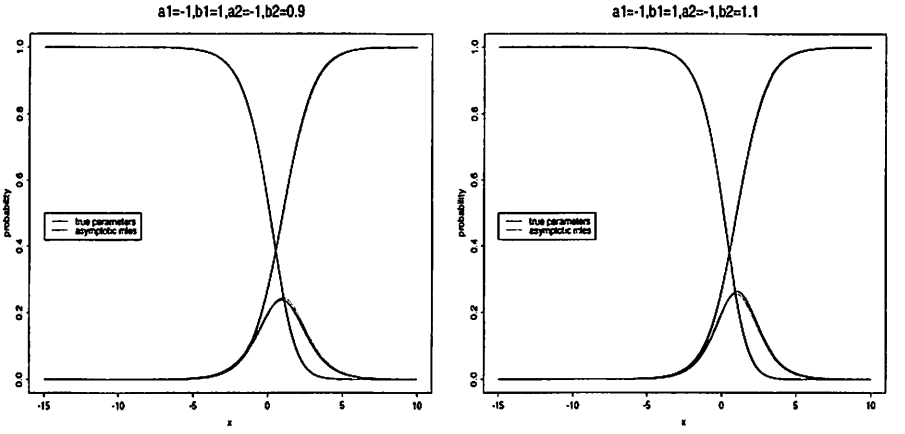


Figure 10: Probability plots, probability vs. dose, for $a_1 = -1, b_1 = 1, a_2 = -1$, and $b_2 = 0.9$ (left); $b_2 = 1.1$ (right). The solid curves are made by the true values of parameters while the dashing curves are made by the asymptotic values of MLE's.

inferences will be explored.

Denote the MLE of x_{max} in the constant-slope model as $\widehat{x}_{max} = -\frac{\widehat{a_1}+\widehat{a_2}}{2\widehat{b_1}}$. This MLE, $\widehat{x}_{max}$, will be used to estimate x_{max}. The asymptotic limit of $\widehat{x}_{max}$, x^*_{max}, is $-\frac{a^*_1+a^*_2}{2b^*_1}$. Two errors are defined:

$$\begin{aligned} error &= \widehat{x}_{max} - x_{max} \\ asyerror &= x^*_{max} - x_{max}. \end{aligned}$$

Also let $A.S.E.$ be the estimated asymptotic standard error of $\widehat{x}_{max}$ in the constant-slope model ($\widehat{a_1}$, $\widehat{b_1}$, and $\widehat{a_2}$ as parameter estimates), and let $A.A.S.E.$ be the limit of $A.S.E.$, which can be obtained by using a^*_1, b^*_1, and a^*_2 as parameters. Let $\widehat{\theta} = (\widehat{a_1}, \widehat{b_1}, \widehat{a_2})$, $\theta^* = (a^*_1, b^*_1, a^*_2)$, and $M(\theta, \eta)$ be as defined in

Appendix A. Then:

$$\begin{aligned} A.S.E. &= \sqrt{c(\widehat{\theta})^T M(\widehat{\theta},\eta)^{-1} c(\widehat{\theta})} \\ A.A.S.E. &= \sqrt{c(\theta^*)^T M(\theta^*,\eta)^{-1} c(\theta^*)} \end{aligned}$$

If a_1, a_2, r are fixed:

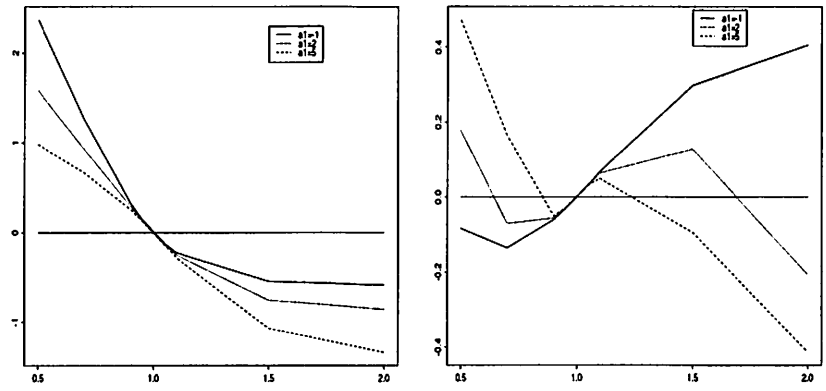
1. *asyerror* is equal to (*asyerror* for $b_1 = 1)/b_1$
2. *A.A.S.E.* is equal to (*A.A.S.E.* for $b_1 = 1)/b_1$

Assume also that the sample size, n, is 100. Figure 11 illustrates how *asyerror* changes as b_2 changes, for different u and a_1 values. The usual approximate 95% confidence interval (C.I.) of x_{max} is $\widehat{x}_{max} \pm 1.96 A.S.E.$. Figure 12 shows the asymptotic 95% C.I., $x^*_{max} \pm 1.96 A.A.S.E.$ for $u = -3$ and different a_1 values.

In Fan's 1999 dissertation many plots for different cases are shown systematically for $0.9 < b_2 < 1.1$ (b_2 is less than 10% different from b_1). In all plots the $|asyerror|$ and *A.A.S.E.* are both small if b_2 is less than 10% different from b_1.

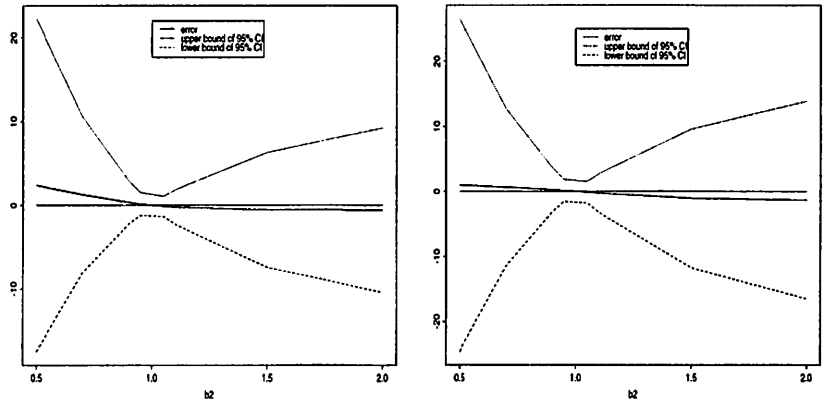
5.3 Small Sample Performance

This section will explore the estimation of x_{max} using c-optimal designs (for estimating x_{max}) and D-optimal designs, for a sample size of 20.



(a) $u = -3$ (b) $u = 5$

Figure 11: Plots of *asyerror* vs. b_2. The solid lines are for $a_1 = -1$, the dashed lines are for $a_1 = 2$, and the double dashed lines are for $a_1 = 5$.

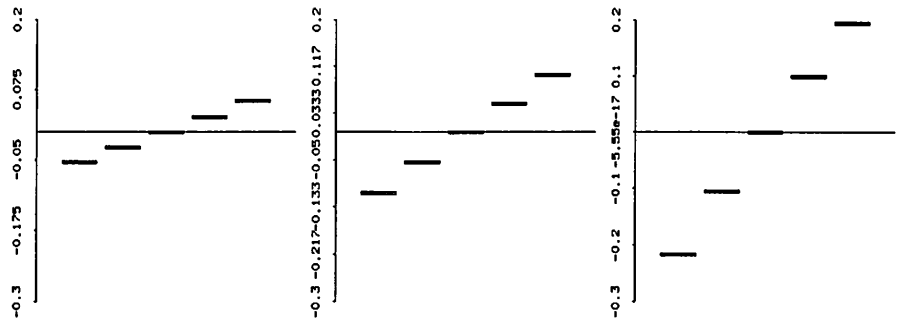


(a) $a_1 = -1$ (b) $a_1 = 5$

Figure 12: Plots of asymptotic 95% C.I. for $u = -3$, and $n = 100$. The solid line is *error*, the dashing line is the upper limit of the 95% C.I., and the double dashing line is the lower limit of the 95% C.I..

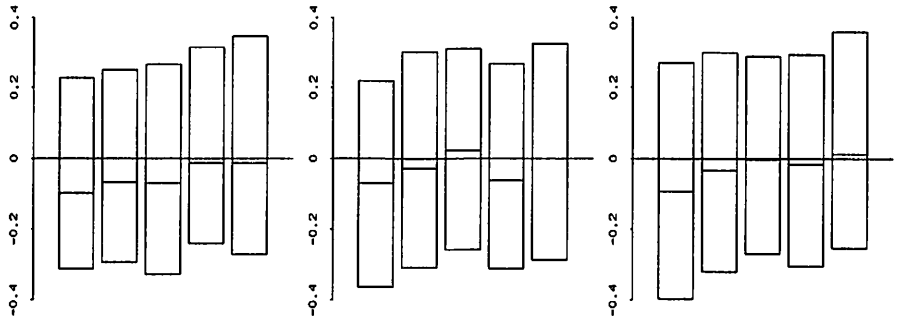
The distribution of error, $\widehat{x}_{max} - x_{max}$, of incorrectly using optimal designs for the constant-slope model with $\widetilde{\theta} = (u, a_1, b_1)$ while a true parameter vector, θ, has the identical u, a_1, and b_1, but b_2 is not equal to b_1, does not depend on the value of b_1. Hence, $b_1 = 1$ is assumed. Since a large value of a_1 is seldom expected, only $a_1 = -1, 0$, and 2 were explored. If $r = b_1/b_2$ is not close to one, the constant-slope model should not be used, so only $r = 0.9, 0.95, 1, 1.05$, and 1.1 were investigated. Recall that if r is close to 1, the value of $u = a_2 - a_1$ indicates how good this drug/therapy is, so several u values: $-1, 0, 1$, and 5 were chosen. To illustrate, boxplots of *error* were drawn.

There were several extremely outlying points in these boxplots of *error*. Hence "box only" plots are shown here, which only show the boxes, not the whiskers. The boxes are bounded by the upper and lower quartiles. Figure 13 shows results for c-optimal designs, and Figure 14 shows results for D-optimal designs. It was found that the interquartile range (IQR) was quite small and *error* was, on average, also small, although the tails were long. In most cases, the IQR of *error* of using D-optimal designs was larger than that of using c-optimal designs. In the $u = -1, a_1 = 0$, and $r = 0.9$ case, however, the IQR of *error* of using D-optimal designs was smaller than that of using c-optimal designs.



(a) $a_1 = -1$ (b) $a_1 = 0$ (c) $a_1 = 2$

Figure 13: "Box only" boxplots of error. In each boxplot, $u = 1$ and $r = 0.9, 0.95, 1, 1.05$, and 1.1 from left to right, using c-optimal designs



(a) $a_1 = -1$ (b) $a_1 = 0$ (c) $a_1 = 2$

Figure 14: "Box only" boxplots of error. In each boxplot, $u = 1$ and $r = 0.9, 0.95, 1, 1.05$, and 1.1 from left to right, using D-optimal designs

5.4 Summary

To summarize, if b_2 is less than 10% different from b_1 and the D-optimal designs of the constant-slope model are used then the designs are very efficient and also, asymptotically, the fitted probability curves are very close to the true probability curves. In addition, when b_2 is less than 10% different from b_1, c-optimal designs for the constant-slope model provide a reasonable estimate of x_{max}, even for a sample size of 20.

The results in this section have generally confirmed Thall's and Russell's (1998) assumption that the constant-slope model performs reasonably well if the different-slope model is true and the slopes are close. Moreover, a useful guideline for how close is "close" has been found: if b_2 is within 10% of b_1, this is close enough. In addition, for a small sample, $n = 20$, the performances of c-optimal designs were good.

6 Conclusions

This paper has presented designs for a trinomial response model. Limiting optimal designs, under the constant slopes assumption, have been introduced and shown to be very efficient. The concept of a sequence of limiting optimal designs is very general and can potentially be applied to other models.

Appendix A: Information Matrices

Based on the trinomial distribution of a single outcome at a dose level x the Fisher information matrix is:

1. for the constant-slope model ($b_1 = b_2$):

$$I(\theta, x) = \left\{ p_3(\theta,x)(p_1(\theta,x)+p_2(\theta,x)) \begin{bmatrix} 1 & x & 0 \\ x & x^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{p_1(\theta,x)p_2(\theta,x)}{p_1(\theta,x)+p_2(\theta,x)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & x^2 & x \\ 0 & x & 1 \end{bmatrix} \right\}$$

   where θ is (a_1, b_1, a_2).

2. for the different-slope model ($b_1 \neq b_2$):

$$I(\theta, x) = \frac{e^{a_2+b_2x}}{(1+e^{a_2+b_2x})^2(1+e^{a_1+b_1x})} \begin{bmatrix} 1 & x & 0 & 0 \\ x & x^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \frac{e^{a_1+b_1x}}{(1+e^{a_1+b_1x})^2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & x & x^2 \end{bmatrix}$$

where $\theta = (a_2, b_2, a_1, b_1)$.

For a design η putting mass m_i at x_i, $i = 1, 2, ..., k$, and $\sum m_i = 1$ the Fisher information matrix is $M(\theta, \eta) = \sum_i \eta_i I(\theta, x_i)$.

Appendix B: General Equivalence Theorem

Theorem. *Define $\mathcal{H}$ to be the set of probability measures on $\mathcal{X}$ and define η_x to be the probability measure putting mass one at $x \in \mathcal{X}$. Suppose a criterion ϕ is concave on $\mathcal{H}$ and differentiable at η^* in $\mathcal{H}$. The following are equivalent*

1. *$\phi(\eta)$ is maximized for $\eta \in \mathcal{H}$ at η^*.*
2. *$sup_{x \in \mathcal{X}} F_\phi(\eta^*, \eta_x, \pi) = 0$.*

For D-optimality and two designs, η_0 and η, and θ a vector of p unknown parameters: (π is a prior distribution of θ.)

$$F_\phi(\eta_0, \eta, \pi) = E_\pi[\text{Trace}\{M(\eta, \theta)M^{-1}(\eta_0, \theta)\}] - p.$$

Similarly for c-optimality:

$$F_\phi(\eta_0, \eta, \pi) = E_\pi[c^T(\theta)M^{-1}(\eta_0, \theta)M(\eta, \theta)M^{-1}(\eta_0, \theta)c(\theta) - c^T(\theta)M^{-1}(\eta_0, \theta)c(\theta)].$$

For $\eta = \eta_x$ these derivatives are simple functions of x.

Appendix C: Proof of Lemma 2.1

Proof. By concavity of ϕ, the following is easily shown, see also Silvey (1980, page 18):

$$\begin{aligned} F_\phi(\eta_i, \eta_i^*, \pi_i) &= \mathrm{E}_{\pi_i}[\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\{\phi(\theta, (1-\epsilon)\eta_i + \epsilon\eta_i^*) - \phi(\theta, \eta_i)\}] \\ &\geq \mathrm{E}_{\pi_i}[\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\{(1-\epsilon)\phi(\theta, \eta_i) + \epsilon\phi(\theta, \eta_i^*) - \phi(\theta, \eta_i)\}] \\ &= \mathrm{E}_{\pi_i}[\phi(\theta, \eta_i^*) - \phi(\theta, \eta_i)] \\ &= \mathrm{E}_{\pi_i}\phi(\theta, \eta_i^*) - \mathrm{E}_{\pi_i}\phi(\theta, \eta_i) \\ &= \phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i). \end{aligned}$$

Therefore,

$$\begin{aligned} \phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i) &\leq F_\phi(\eta_i, \eta_i^*, \pi_i) \\ &\leq \sup_\eta F_\phi(\eta_i, \eta, \pi_i) \\ &= d_i. \end{aligned}$$

Since $d_i \to 0$ as $i \to \infty$, $\phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i)$ goes down to zero as $i \to \infty$. □

References

Agresti, A. (1990). *Categorical Data Analysis.* John Wiley & Sons, New York.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and In-*

ference **21**, 191-208.

Clyde, M. (1993). An object-oriented system for Bayesian nonlinear design using Xlisp-Stat. *Technical Report #587, School of Statistics, University of Minnesota.*

Fan, S. (1999). *Multivariate Optimal Designs.* PhD Thesis, School of Statistics, University of Minnesota. http://www.stat.umn.edu/~ fan/thesis.html

Ford, I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Jour. Roy. Statist. Soc., Ser. B* **54**, 569-583.

Heise, M. A. and Myers, R. H. (1996). Optimal designs for bivariate logistic regression. *Biometrics* **52**, 613-624.

Glonek, G. F. V. (1996). A class of regression models for multivariate categorical responses. *Biometrika* **83**, 15-28.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* **57**, 533-546.

Hoffman, K. and Kunze, R. (1971). *Linear Algebra.* Englewood Cliffs: Prentice-Hall.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium.* Berkeley: University of California Press, 221-233.

Kitsos, C.P., Titterington, D.M. and Torsney, B. (1988). An optimal design problem in rhythmometry. *Biometrics* **44**, 657-671.

Krafft, O. and Schaefer, M. (1992). D-optimal designs for a multivariate regression model. *Journal of Multivariate Analysis* **42**, 30-140.

Morrison, D. (1976). *Multivariate Statistical Methods.* McGraw-Hill.

Thall, P. F. and Russell, K. E. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* **54**, 251-264.

Thall, F. P. , Estey, E. H. , and Sung, H. G. (1999). A new statistical method for dose-finding based on efficacy and toxicity in early phase clinical trials. *Investigational New Drugs* **17**, 155-167.

Silvey, D. S. (1978). Optimal design measures with singular information matrices. *Biometrika* **65**, 553-559.

Silvey, D. S. (1980). *Optimal Design.* Chapman and Hall.

Sitter, R.R. and Wu, C.F.J. (1999). Two-stage design of quantal response studies. *Biometrics* **55**, 396-402.

Verdinelli, I. (1992). Advances in Bayesian experimental design. In *Bayesian Statistics 4*, J.M. Bernardo et al eds., Clarendon Press, Oxford, 467-475.

White, L. V. (1975). *The Optimal Design of Experiments for Estimation in Nonlinear Models.* PhD Thesis, University of London.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society, Series B* **35, No. 1**, 123-130.

Wong, W.K. and Lachenbruch, P.A. (1996). Designing studies for dose response. *Statistics in Medicine* **15**, 343–359.

Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. *Optimal Design and Analysis of Experiments.* Y. Dodge et al, eds. North-Holland, Amsterdam, 213- 224.

Zhu, Y. , Krewski, D. , and Ross, W. H. (1994). Dose-response models for correlated multinomial data from developmental toxicity studies. *Applied Statistics* **43**, 583-598.

Zocchi, S. S. and Atkinson, A. C. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics* **55**, 437-444.